ISSN: 0975-766X
CODEN: IJPTFI
Available Online through ***Research Article***
www.ijptonline.com

DEVELOPMENT AND RESEARCH OF STATISTICAL METHODS AND OPTIMIZATION ALGORITHMS OF SEARCH FOR SOLUTIONS IN INTELLIGENCE AUTOMATED SYSTEMS

V.A. Atroshchenko*, V.Ye. Belchenko, I.V. Belchenko, R.A. Dyachenko
Armavir State Pedagogical University, 159 Rosa Luxemburg Street, Armavir.

Received on 03-06-2016 *Accepted on 27-06-2016*

Abstract:

The modern approach to the automation of various fields of human activity involves intellectualization of automated control systems. The most common types of intelligence systems are expert systems. As a result of increasing volumes of knowledge bases it became necessary to prepare methods that accelerate the decision-making procedure by expert systems.

The basic idea of optimization of search for solutions is to minimize the questions asked by the user, and hence increase the speed of search for solutions.

The research deals with the development of optimization methods of search for solutions in the rule-based expert systems. A method of optimizing the search for solutions in the rule-based expert systems by ordering hypotheses in descending mathematical expectation of their degree of reliability has been prepared. A method for automatic structuring of the set of hypotheses based on the study of correlation dependence between them has been developed. A method for evaluating the effectiveness of a production expert systems based on the use of a simulation model of an expert system has been suggested. The architecture of the simulation expert system has been developed and the study algorithm of methods optimization efficiency of search for solutions of the model has been proposed. The results of the study of effectiveness of the proposed optimization methods in the dissertation have been obtained.

Keywords: Optimization, Rule-based expert systems, Automatic structuring of set of hypotheses

1. Introduction

This paper develops and investigates the methods that are based on accounting the statistics on the results of search for solution in the rule-based expert system. As a result of search in the rule-based expert system with fuzzy logic, one calculates the degrees of reliability of the hypotheses to determine the true hypothesis. Such statistical population can be the subject of analysis for the construction of a meta-level and factorization of the search space.The paper develops and investigates two optimization procedures of search: the factorization method of search space by considering the correlation of degrees of reliability of the hypotheses and the method of ordering in the process of proving descending mathematical expectation of their degree of reliability.

2. The Development Of Statistical Methods For Optimization Of Retrieval In Intelligence Automated Systems

Let $H = \{h_1, h_2, \ldots, h_n\}$ is a set of hypotheses in the production system. We denote the degree of reliability of hypothesis h_j by μ_j. As a result of the output system operation in the expert system a corresponding set of hypotheses of degrees of reliability $(\mu_1, \mu_2, \ldots, \mu_n)$ has been obtained. We assume that N implementations of solutions search procedure in the expert system have been conducted. For each of the realizations, the corresponding sets of the degree of reliability of hypotheses have been obtained

$$(\mu_{1i}, \mu_{2i}, \ldots, \mu_{ni}), \quad i = 1, 2, \ldots, N.$$

The set computed in this way provides statistical estimation of the degrees of hypotheses reliability as mathematical expectation. Mathematical expectation of degree of reliability of i hypothesis that meets the requirements of consistency and unbiasedness has the appearance[1,2,3]

$$m_i = \frac{1}{N} \sum_{k=1}^{N} \mu_{ik}$$

The set of mathematical expectations of the degrees of hypotheses reliabilities $(m_1, m_2, \ldots, m_n)$ allows to order a hypotheses sequence $h_1, h_2, \ldots, h_n$ in the form of $h_l, h_m, \ldots, h_k$ so that $m_l \geq m_m \geq \ldots \geq m_k$. This ordering makes it possible to prove the hypotheses with maximum expectation value of the degree of reliability first.

The basic computational procedures in the expert system associated with a set of statistics are made during expert system operation.The sets of degrees of reliability of the hypotheses obtained in the process of operation $(\mu_{1i}, \mu_{2i}, \ldots, \mu_{ni}), \quad i = 1, 2, \ldots, N.$ are accumulated and stored in computer memory. Initially, an arbitrary ordering in the form of hypotheses $h_1, h_2, \ldots, h_n$ is selected. Computation of mathematical expectations of the degrees of hypotheses reliability and hypotheses ordering are performed at the stage of preparing a system to operation, and the resulting sequence $h_l, h_m, \ldots, h_k$, where $m_l \geq m_m \geq \ldots \geq m_k$, are stored in computer memory.

The advantages of the method of ordering hypotheses descending the degrees of mathematical expectations of the degrees of hypotheses reliability include:

1) independence of the analysis on the form of the function of estimation of degree of reliability (these functions can be polynomial, exponential, etc.);

2) possibility of re-analysis when changing the base of knowledge;

3) dynamic ordering of hypotheses in the process of operation.

3. The Method of Automated Structurization of The Set of Hypotheses Based on the Research of Correlation Between Them

Let $F=\{f_1, f_2,..., f_m\}$ is asset of terminal facts, and $H=\{h_1, h_2,..., h_n\}$ -a set of hypotheses in the production system. Let ν_i denote the degree of reliability of the fact f_i, and μ_j - the degree of reliability of hypothesis h_j. Suppose that the value set $(\nu_1, \nu_2,..., \nu_m)$ of the degree of reliability of terminal facts comes in the input of the production system. As a result of output system operation in the expert system, a corresponding set of the degrees of hypotheses reliability $(\mu_1, \mu_2,..., \mu_n)$ is obtained. Now we suppose that the set $(\nu_1, \nu_2,..., \nu_m)$ is a specific dimensional m uniformly distributed in hypercube $[0,1]^m$ of random variable and N realizations of this variable have been performed. For each of the realizations, the proper sets of degree of reliability of hypotheses $(\mu_{1i}, \mu_{2i},..., \mu_{ni})$, $i=1,2,...,N.$ have been obtained.

This statistical population enables to obtain the estimations of basic numerical characteristics of hypotheses. The mathematical expectation and the variance of the degree of reliability of i hypothesis that meets the requirements of consistency and unbiasedness have the form [4.5]

$$m_i = \frac{1}{N}\sum_{k=1}^{N}\mu_{ik}; \quad D_i = \sigma_i^2 = \frac{1}{N-1}\sum_{k=1}^{N}(\mu_{ik} - m_i)^2$$

The correlation between the hypotheses h_i and h_j is determined by the formula [2,6]

$$K_{ij} = \frac{1}{N-1}\sum_{k=1}^{N}(\mu_{ik} - m_i)\,(\mu_{jk} - m_j)$$

Since we are interested mainly in the degree of correlation between the hypotheses independently of dispersions of their degree of reliability, we will consider the correlation coefficients[2,7] of the hypotheses h_i and h_j and

$$r_{ij} = \frac{K_{ij}}{\sigma_i \sigma_j}$$

The correlation coefficient is selected as a measure of hypotheses proximity. This proximity cannot always, of course, reflect the hypotheses semantics, but it reflects the relationship of their degrees of reliability.

The set of correlation coefficients $r_{ij}, i, j = 1, 2, ..., n$ forms normalized correlation matrix $R = \|r_{ij}\|_{i,j=1}^{n}$.

Asisgenerally known [2], the correlation coefficients satisfy the condition $-1 \le r_{ij} \le 1$ .At that, the greater the module $|r_{ij}|$ is, the stronger linear dependence between the degrees of reliability of the hypotheses h_i and h_j. The coefficient sign r_{ij} points to the presence of positive or relative correlation. Consider the matrix $\bar{R} = \|r_{ij}\|_{i,j=1}^{n}$. The matrix defined in this way induces fuzzy relation $\Re$ to the set of hypotheses H, characterizing the degree of pair dependence of the hypotheses.

Statement. The relation $\Re$ is a fuzzy relation of tolerance.

Proof. As follows from the properties of the coefficients of correlation, $|r_{ii}| = I, |r_{ji}| = |r_{ij}|, i, j = 1, 2, ..., n,$ than the relation $\Re$ is reflexive and symmetric. We show that more generally, $\Re$ is not transitive. Sufficeit only to consider an example in Figure 1.

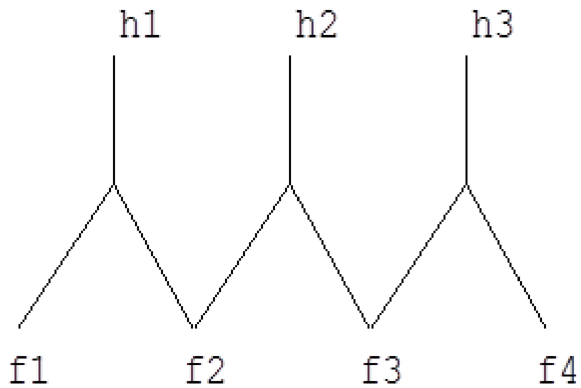


Fig. (1). – An example of the structure of knowledge base.

The structure of the rules in the figure is that $|r_{12}|,|r_{23}|>0$ as the hypotheses of each of the pairs $\{h_1,h_2\}$ и $\{h_2,h_3\}$ have common terminal facts (correspondingly f_2 and f_3), which will condition the dependence of these hypotheses. On account of the similar points $|r_{13}|=0$. At that an inequality required for transitivity [8] $|r_{13}|\geq\min\{|r_{12}|,|r_{23}|\}$ in this case, probably, is not satisfied.

Set at ask to divide the range of hypotheses H in to non-overlapping classes so that each class to represent a group of closely related hypotheses. To immediately doat ask using the relation $\Re$ will be abortive, as it creates subdividing into non-overlapping classes (since according to the statement, $\Re$ - tolerance relation). To find subdividing into disjoint classes it is necessary $\Re$ to be converted into equivalence relation $\hat{\Re}$, which can be fulfilled using a transitive closure procedure [9].

A characteristic feature of fuzzy relation is that by partitioning α-levels from $\hat{\Re}$ one can get a family of normal relations which will be corresponded by a family of nested set partitions of $H:\{H_{\alpha i}, i=0,1,...,k;\alpha_0=0,\alpha_k=1\}$,

$H_1\subset H_{\alpha k-1}\subset...\subset H_{\alpha 1}\subset H_0$, где H_1 и H_0 - transitive partitions

(Figure 2).

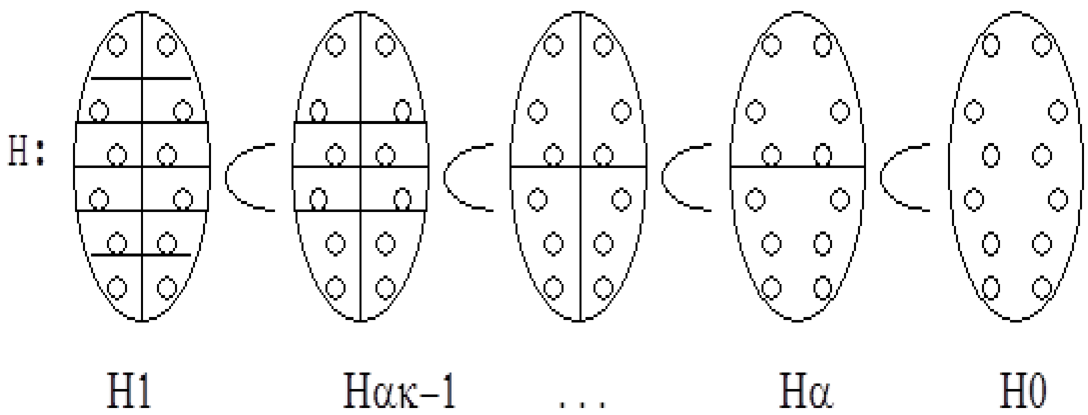


Fig. (2). – An example of hypotheses set partitioning into α – levels.

Consider a partition $H_{\alpha i}: H_1 \cup H_2 \cup ... \cup H_0 = H_1$, $H_i \cap H_j = 0$ $i, j = 1,2,...,n,$

$i \neq j;\ H_s = \{h_{st}, t = 1,2,...,n_s\},\ s = 1,2,...,p, n_1 + n_2 + ... + n_p = n = |H|.$

Select one of the hypotheses in each of the classes, for example, h_{s1}, and by using top-down search procedure, find the degree of reliability μ_{s1} of each of them. The computed values $\mu_{s1}, s = 1,2,...,P$ allow with in each class to calculate an expected degree of reliability of the rest of hypotheses. To use a line arregress ion equation is sufficient for that. These equations for class $H_s, s = 1,2,...,P$ have the form of

$$\mu_{st} = m_{st} + \frac{\sigma_{st}}{\sigma_{s1}} r(h_{st}, h_{s1})(\mu_{s1} - m_{s1}), t = 2,3,..,n_s$$,‹ где $r(h_{st}, h_{s1}) -$

correlation coefficient between the hypotheses h_{st} и h_{s1}. The greater $|r(h_{st}, h_{s1})|$ is, the more exact the value of the degree of reliability μ_{st}. predicted by the formula will be. Computed in this way degrees of reliability enable to order all the hypotheses in knowledge data descending their preliminary degrees of reliability [15]. It is obvious that this ordering determine at the same time an optimal sequence of the hypotheses being proved in the expert system. In other words, one should first prove the hypothesis with maximum suggested degree of reliability, as most probably it is true, then the hypothesis following it and so on.

Theprocessoftruehypothesissearchcanbestoppedatanytime,rememberingthatthe hypothesis that is more likely to be true is one that has long since been already proven (that has the greatest degree of reliability), or on reaching the hypothesis of the degree of reliability which exceeds a predetermined threshold value. It must be emphasized that since the procedure of finding the degree of reliability of the hypothesis takes immeasurably more time in case of output in knowledge base than the calculated by regression formulas, then the found ordering of hypotheses can dramatically break the expert system work short. As is generally known, slow system operation and a large number of questions asked are disadvantages of the rule-based expert systems [16]. Basic computational procedures in the expert system associated with a set of statistics, and estimation of point characteristics of hypotheses and construction of normalized correlation matrix, are performed at preparation stage for operation of the system, and then stored in computer memory for the study during its operating. It should be noted that the uniform distribution of the degrees of reliability of the terminal facts at the set of the hypotheses statistics may not correspond to the real distribution.

More or less definite answer to this question can only be received only in the course of operation of the expert system. Inthisconnectionitisrationalintheprocessoftheexpertsystemoperationtocarry on statistics of distribution of the degrees of reliability of terminal facts so as to do further adequate research of correlation dependence of hypotheses [17].

In conclusion, it should be noted that the method mentioned above is based on allowing for only the linear dependence of hypotheses. If this relation is more complex, than the corresponding direct regressions are linear approximations of the true regression lines. In this connection, there occurs the task of fixing numerical characteristics that allow to find out the existence of dependence of a higher order than the linear one on the basis of statistical sampling and restore the degree of reliability of hypotheses by means of regression curves [10].

The advantages of the method of investigation of correlation dependence of hypotheses include:

1) independence of analysis on the form of function of estimating the degree of reliability (these functions can be polynomial, exponential, etc.);

2) possibility of re-analysis when changing the volume of knowledge base;

3) presence of several different partitions, which ensures a certain freedom in selection of abstraction level according to hypotheses of the classes. Revelation of the classes of interrelated hypotheses that are likely to be unknown for the experts and be the subject of research In the area of expertise.

4. Experimental Study of The Methods For Optimization Of Solution Search

The studies were conducted with the following three methods of selecting hypotheses to prove:

- Linear search method is in successive selecting the hypotheses to prove. This approach is realized in the majority of the production expert systems;

- Method for ordering a list of hypotheses to select on the basis of reducing the mathematical expectation of their degree of reliability. This approach has been developed and suggested in this work;

- Method of automatic structuring the set of hypotheses on the basis of the analysis of the coefficients of degrees of reliability of hypotheses and dynamic prediction of reliability of the unproved hypotheses through reliability of the proven ones. This method developed in the work is the most effective among the proposed to study.

The effectiveness of the technology such as the application of met a rules is not investigated in this work, since this method uses the expertise knowledge to build the meta-level, depends on a particular subject area, specific expert system and is not a universal method for optimization of search for solutions [12].

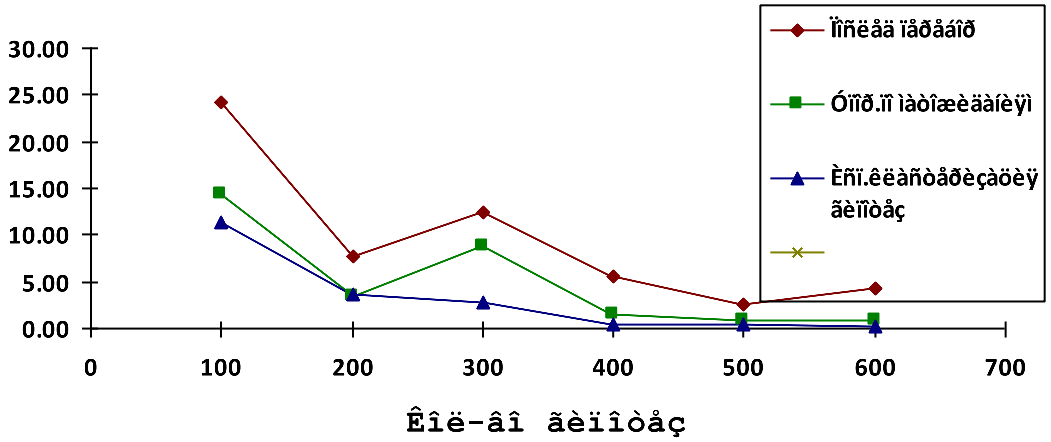
All three methods are investigated on the same knowledge bases for a comparative analysis of the developed methods. Estimate of effectiveness has been performed with the above mentioned parameters such as the number of hypotheses, the number of levels, the maximum number of terms in a rule. In addition, an integral performance indicator has been defined. The functions such as minimum, maximum and mean as the most common in existing systems have been equally probable set up as the membership functions for processing fuzzy knowledge. All the property values are determined by the real results of the experiments.

5. The Study Of Dependence Of Effectiveness Of Methods For Opptimization Of Search For Solutions On The Number Of Hypotheses In Knowledge Base

Figure 3 shows the results of studies of efficacy of the methods for search optimization depending on the number of hypotheses in the knowledge base. The parameter of the number of levels is constant and equal to 4. The measurement result is a relative number of the proven hypotheses (the ratio of number of proven hypotheses to the total number of hypotheses in the knowledge base).

Кол-во гипотез	100	200	300	400	500	600
Послед перебор	24,30	7,68	12,45	5,64	2,54	4,34
Упор.по матожиданиям	14,30	3,34	8,75	1,52	0,8	0,95
Исп.кластеризация гипотез	11,27	3,70	2,78	0,40	0,42	0,27

The number of hypotheses		200	300	400	500	600
Sequential selection	24,30	7,68	12,45	5,64	2,54	4,34
Mathexpectationordering	14,30	3,34	8, 75	1,52	0,8	0,95
Clusterization of hypotheses	11,27	3,70	2,78	0,40	0,42	0,27



The number of hypotheses

Fig. (3). –The table of experimental data and the diagram of dependence of the relative number of proven hypotheses on the total number of hypotheses in the knowledge base.

6. The Study Of Dependence Of Effetivenes Of Methods For Optization For Maximum Number Of Conditions In The Rules Of Knowledge Base

The following (Figure 4) results of the experiments characterize the efficacy of the methods for optimization of retrieval depending on the number of levels in knowledge base.

Кол-во условий	1	2	3	4
Послед перебор	4,73	6,71	8,5	9,4
Упор.по матожиданиям	4,03	4,91	6,7	7,09
Исп.кластеризация гипотез	2,70	3,41	3,02	3,72

The number of hypotheses		2	3	4
Sequential selection	4,73	6,71	8,5	9,4
Mathexpectationordering	4,03	4,91	6, 7	7,09
Clusterization of hypotheses	2,70	3,41	3,02	3,72

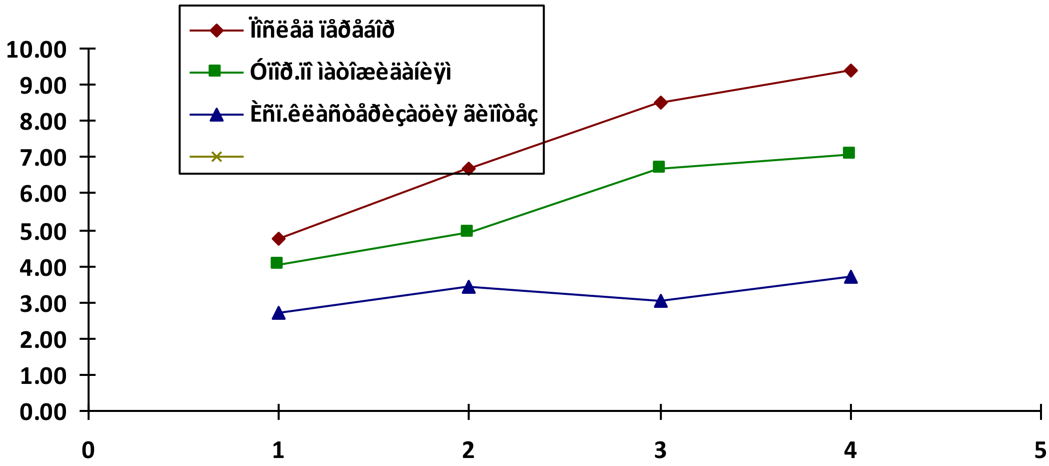


Fig. (4). –The table of experimental data and the diagram of dependence of the relative number of proven hypotheses on number of conditions in knowledge base

The results indicate that the method of automatic factorization of search space for finding solutions requires the method of sequential selection to be more efficient by 63% than the method of ordering expectation by 38%, respectively.

7. Summarization of The Research Results of The Methods For Optimization of Search For Solutions

As the results of the studies described in the previous chapter have shown, the most effective optimization method of search for solution is the method of automatic structuring of the set of hypotheses based on the correlation dependence between them. The field of application of the method can be expert systems design with large databases in terms of knowledge. However, judging by the results of the experiments, in expert systems with bases of knowledge having minor number of hypotheses, the suggested method of ordering hypotheses to reduce the mathematical expectation of their degrees of reliability can be applied in this paper [13]. Such substitute in the systems with small knowledge bases may be more preferable due to the fact that the math expectation ordering method requires much less computational resources than the one discussed above. The field of application is expert systems with small volumes of knowledge base, the advising systems.

A widely used method of sequential selection is appropriate to be applied in the systems with a limitation of computational resources such as hard drives that allow to quickly accumulate statistics. The range of application is small expert systems for home computers [14].

The suggested methods are evolutionary according the criteria of analysis of history of rule-based expert system operation and present state of the process of search for solutions.

The stages of evolutionary [11] development of the prepared optimization techniques to find solutions may be presented in the table in Figure 5.

The method of optimization of search for solution	Analysis of prehistory of expert systems operation	Analysis of the current search
Sequential selection	-	-
Ordering hypotheses of descending mathematical expectation of the degree of reliability	+	-
Automatic structuring of the set of hypotheses based on the study of correlation dependence between them	+	+

Evolutionary stages of development of search optimization methods in the rule-based expert systems.

8. Summary

As a result of experimental research of optimization methods, it can be concluded that the methods proposed in the paper are really more effective than the existing ones and enable to use them in the rule-based expert systems with arbitrary semantics of knowledge base.

These results indicate that the automatic factorization method of search space for finding of solutions requires the method to be more efficient by 63% than the method of sequential selection and the method of ordering expectation - by 38%, respectively.

9. Conclusion

The proposed methods of optimization of solutions search process in intelligence automated systems are versatile in comparison with the existing ones as they do not focus on a specific subject area. The scope of application of these methods are production expert systems with crisp and fuzzy logic. The method of ordering the hypothesis on descending expectations of the degrees of reliability can be effective in expert systems with a small volume of knowledge data base, since it is very easy to be realized and does not require complex computational procedures in the search for solutions. This approach is convenient to use in modifying the shell of the existing expert system.

The method of automatic structuring of the set of hypotheses is based on the study of correlation dependence between them. This approach is effective when dealing with knowledge bases of a significant volume (in awidely used expert systems). Factorization of the set of hypotheses is done on the basis of a correlation dependence between the degrees of reliability of the hypotheses. Using the transitive closure operation for the construction of equivalence relations on a set of hypotheses leads, of course, to some of the correlation coefficients, but it is insignificant approximation compared to the method efficiency.

Both of the proposed methods of optimization automatically adapt not only to a specific knowledge base, but also a specific class of tasks. That is, the method of ordering or the classes of equivalence with the same knowledge base will be different if the expert system is mainly used for various tasks. This distinguishes the proposed method from the available optimization methods of search for solutions currently used in expert systems.

References

1. Manualon Probability Theoryand Mathematical Statistics / V.S. Korolyuk, N.I. Portenko, A.V.Skorokhod, Turbin.-M.: Science. Chief Editorial Board of Physicomathematical Literature, 1985.-640p.
2. BalashovYe.P. Evolution Synthesis of Systems . - M.:Radio communication, 1985, 328 p., il.

3. VentselYe. S., OvcharovL.A. ProbabilityTheoryandits Engineering Applications.-M.: Science. Chief Editorial Board of Physicomathematical Literature-(Physicomathematical Library of an Engineer).-480 p.
4. Statistical Technology for Computer /Edited by K.Enslayn,E. Ralston, G.S. Wilf: Translated from English - M.:Science. Chief Editorial Board of Physicomathematical Literature, 1986.-464p.
5. KolemayevV.A. andothers. Probability Theoryand Mathematical Statistics: Tutorial for higher schools of economics / V.A. Kolemayev,O.V.Staroverov, V.B.Turundayevsky; Edited by V.A. Kolemayev. - M.: Higher School, 1991. - 400 p.
6. GranovskyV.A., SirayaT.N. MethodsofExperimentalDataProcessingby Measuring.-L.: EnergyAtomPress. TheLeningraddepartment, 1990.-288 p.
7. PitmanE. ThefundamentalsofTheoryofStatisticalInference: TranslatedfromEnglish-M.: Mir, 1986.-104 p.
8. Pollard J. A Handbook of Numerical and Statistical Techniques /Translated from EnglishbyV.S. Zanadvorov; Edition and prefaceby Ye. M. Chetyrkin.-M.: FinancesandStatistics, 1982.-344 p.
9. Belchenko V. Ye.,Karishchenko A. N. On a Optimization Method for Search in Expert Systems.- The Collected Methods of Algorithm Modeling in Complex Systems: Interdepartmental Subject Science Book.- Taganrog: TRTI, 1988, Issue 7.-P.58-64.
10. FuzzySetsintheModelsofControlandArtificialIntelligence/EditedbyD.A. Pospelov.-M.: Science. Chief Editorial Board of Physicomathematical Literature, 1986.-312p.- (Problems of Artificial Intelligence).
11. IvakhnenkoA. G., YurachkovskyYu. P. ModelingofComplexSystemson Experimental Data.-M.:Radio communication, 1987.-120p.:il.-(Cybernetics).
12. AtroshchenkoV.A., BelchenkoV.Ye., BelchenkoI.V., DyachenkoR.A.Methods of Optimization Of Network Structure of Technological Network of Electric Power Cluster of TheRegion // International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 21 (2015) pp 41761-42766
13. KonovalovD. P., DyachenkoR.A., Bogdanov V.V. The Advanced Development Tools of Web-Applications. ComparativeAnalysis.// TheCollectionoftheThirdInternationalTheoretical and Practical ofYoung Scientist devoted to the 52-th Anniversary of Yu. A. Gagarin's Flight into Space. 2013. P. 303-306.
14. FisherA.V., Dyachenko R.A., Loba I. S. TheSetupofChronological Data Storage in Data Bases of Monitoring and Forecasting Systems// Multidisciplinary Network Electronic Science Journal of Kuban State Agrarian University. 2012/ № 79. P. 271-280.

15.Sharovatov A.S., LobaI.S., Reshetnyak M.G. Algorithm Design For Search for Optimal Model// Multidisciplinary Network Electronic Science Journal ofKubanState Agrarian University. 2012. № 77. P. 413-422.

16.Kucher V.A., MagomadovA. S., Chiglikova N. D., DyachenkoR.A. Information Security ofComputer Network BasedonIntelligence Systems// Multidisciplinary Network Electronic Science Journal of Kuban State Agrarian University. 2015. № 110. P. 1811-1816.

17.Atroshchenko V.A., RudenkoM.V., DyachenkoR.A., Bagdasaryan R.Kh. To the Problem of Setup of Data Storage in Mobile Application // ScientificWorksofKubanStateTechnologicalUniversity. 2014. № 1. P. 189-197.

Corresponding Author:

V.A. Atroshchenko*,